

FROM XML TO MARC

RDF

BEHIND THE SCENES



ABES = French national agency for HE libraries



www.sudoc.abes.fr



www.idref.fr



www.theses.fr



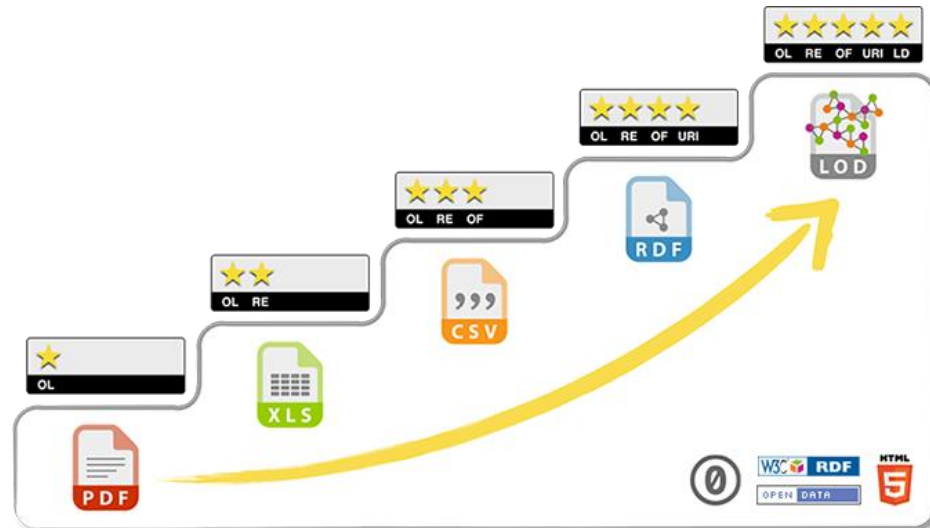
www.calames.abes.fr

General union catalog

Authority data

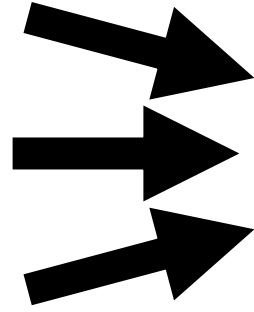
Dissertations (completed or in progress)

Archives and manuscripts

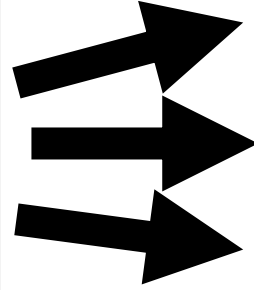


Publisher metadata

A
B
C



RDF



X
Y
Z

marc



marc



marc

Catalogues,
LOD, disco tools ...

Catalogues

COLLECTING

NORMALIZING

SHARING

XML

- ONIX
- JATS
- Etc.

KBART

Etc.



MARC

sruDoc

LOD

Not yet #grrr

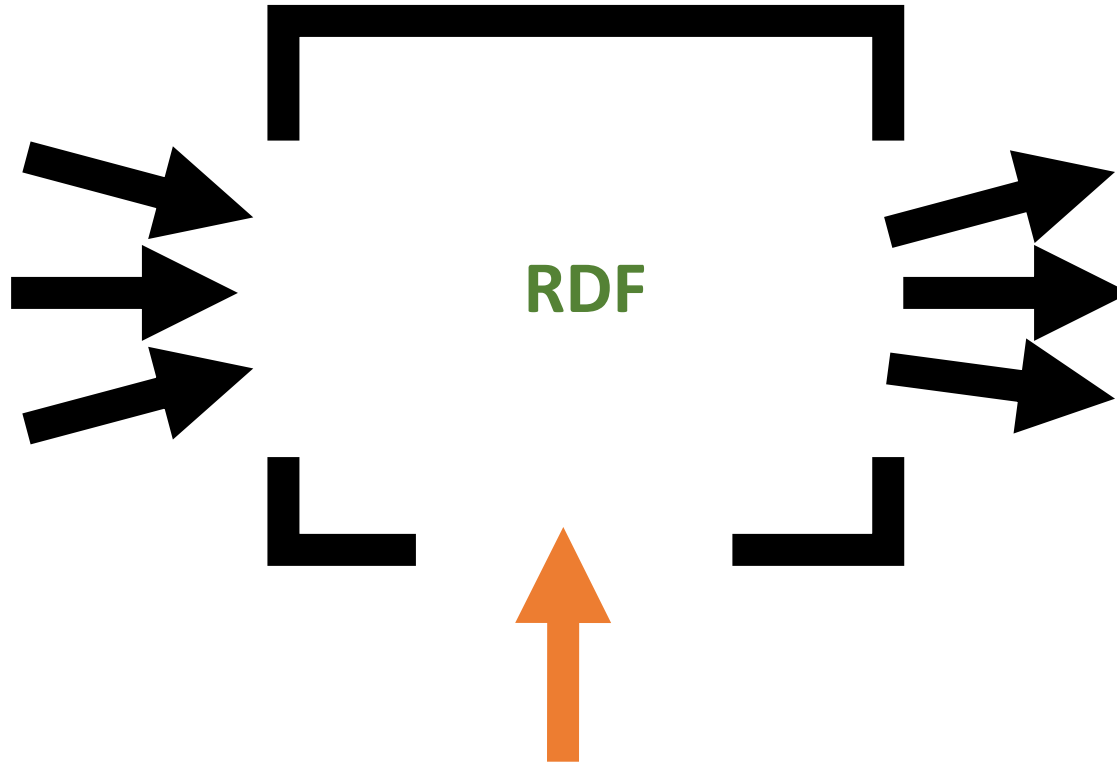
Etc.

XML

- ONIX
- JATS
- Etc.

KBART

Etc.



MARC

LOD

Etc.

ENRICHING

Other bib. Data + reference data (authority data...)

WHAT WE DO

Collecting, Normalizing, Enriching, Sharing

WHAT FOR

Contributing to the **web of open and linked and high quality data**
of scientific content



ON WHICH STUFF

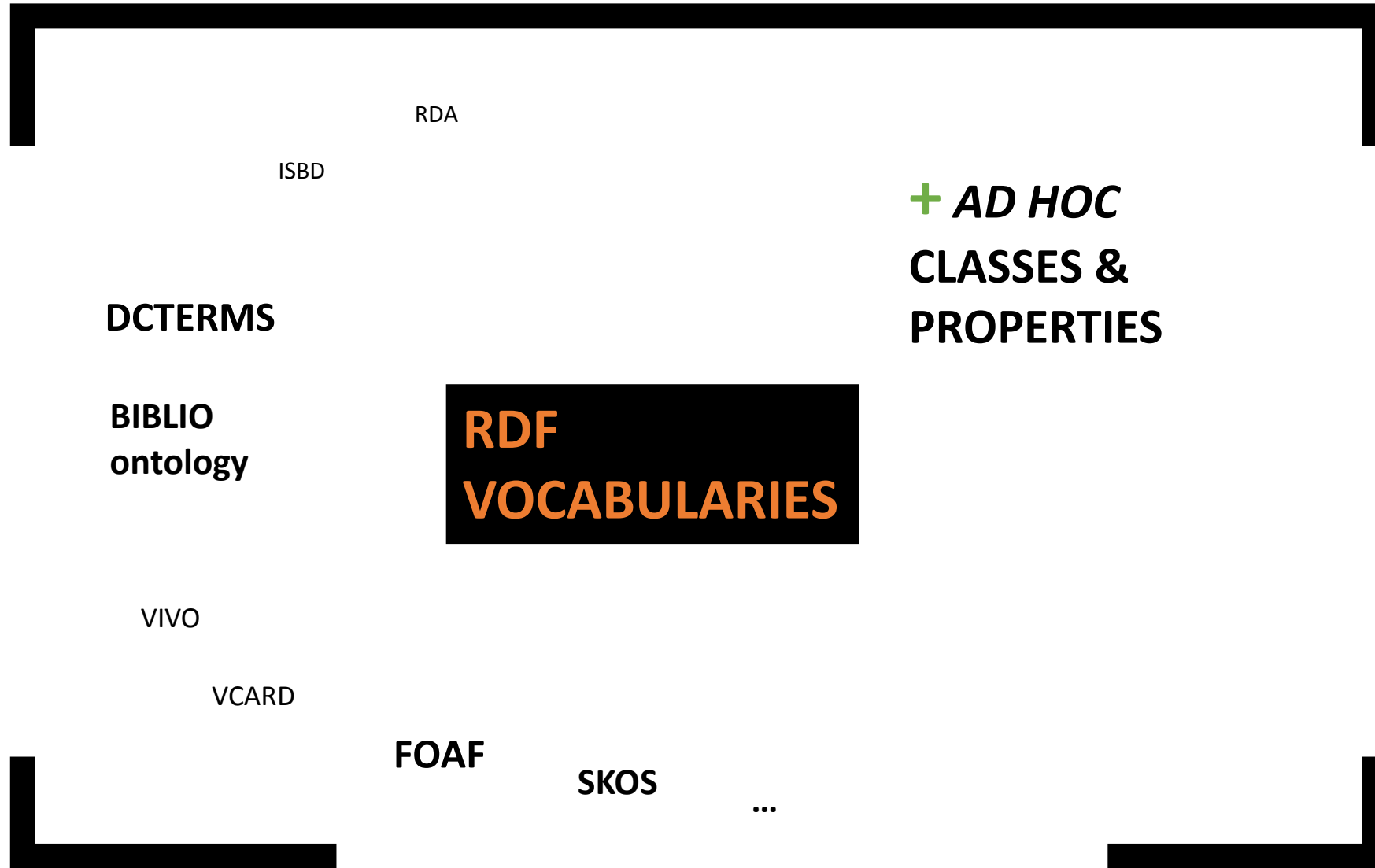
France related metadata sets
+ What is not (yet) in the LOD

We need to be flexible.

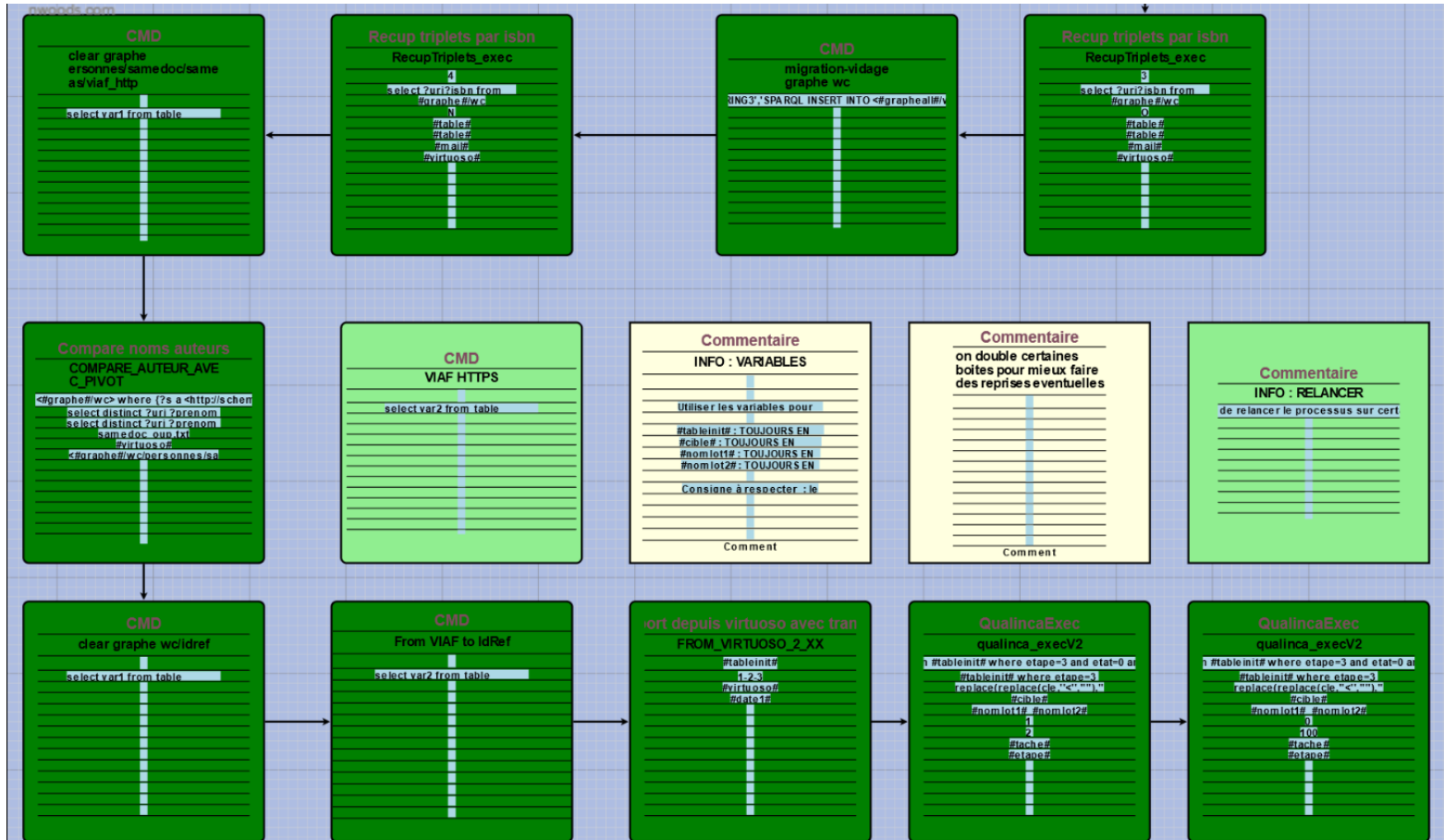
Data model

Workflows

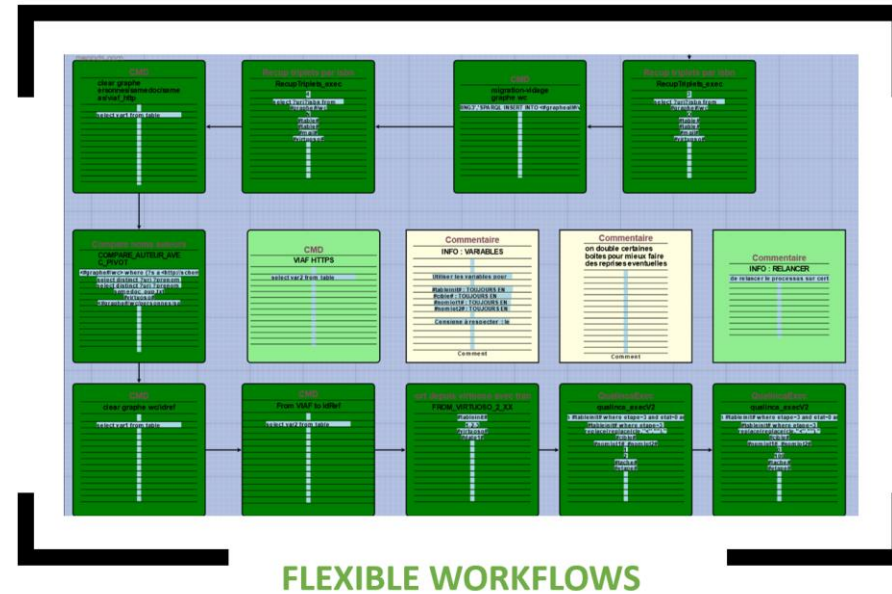
Technological choices



EXTENSIBLE DATA MODEL



FLEXIBLE WORKFLOWS



Basic operations on data > SQL & SPARQL Queries + Stored procedures

Workflow engine > Stored procedures in a relational database

Administration interface for data experts >



ORACLE®

ORACLE®

ORACLE®

ORACLE®



ORACLE®

ORACLE®

TECHNOLOGICAL CHOICES

WHY ORACLE ?

- Need to program close to the data (for performance reasons)
 - PL SQL
 - Java inside Oracle
- Our dependencies
 - Not being dependent on a specific RDF DB ➔ no Virtuoso PL SQL
 - We are already dependent on ORACLE... (and have the skills)
- Joining with more data already in Oracle
 - Mix of SPARQL and SQL (or XQUERY)
- Management of errors ➔ logs in tables, easy to query

4 use cases

XML

RDF

MARC

<OUP_ONIX/>
about each
print book

work

print manifestation

electronic manifestation

MARC
for
print

MARC
for
ebook

print isbn is the
key to merge
the two graphs

From SQL to RDF

| print isbn | elec. isbn | title | date | url |
|------------|------------|-------|------|-----|
| | | | | |
| | | | | |

ORACLE®

KBART about ebooks

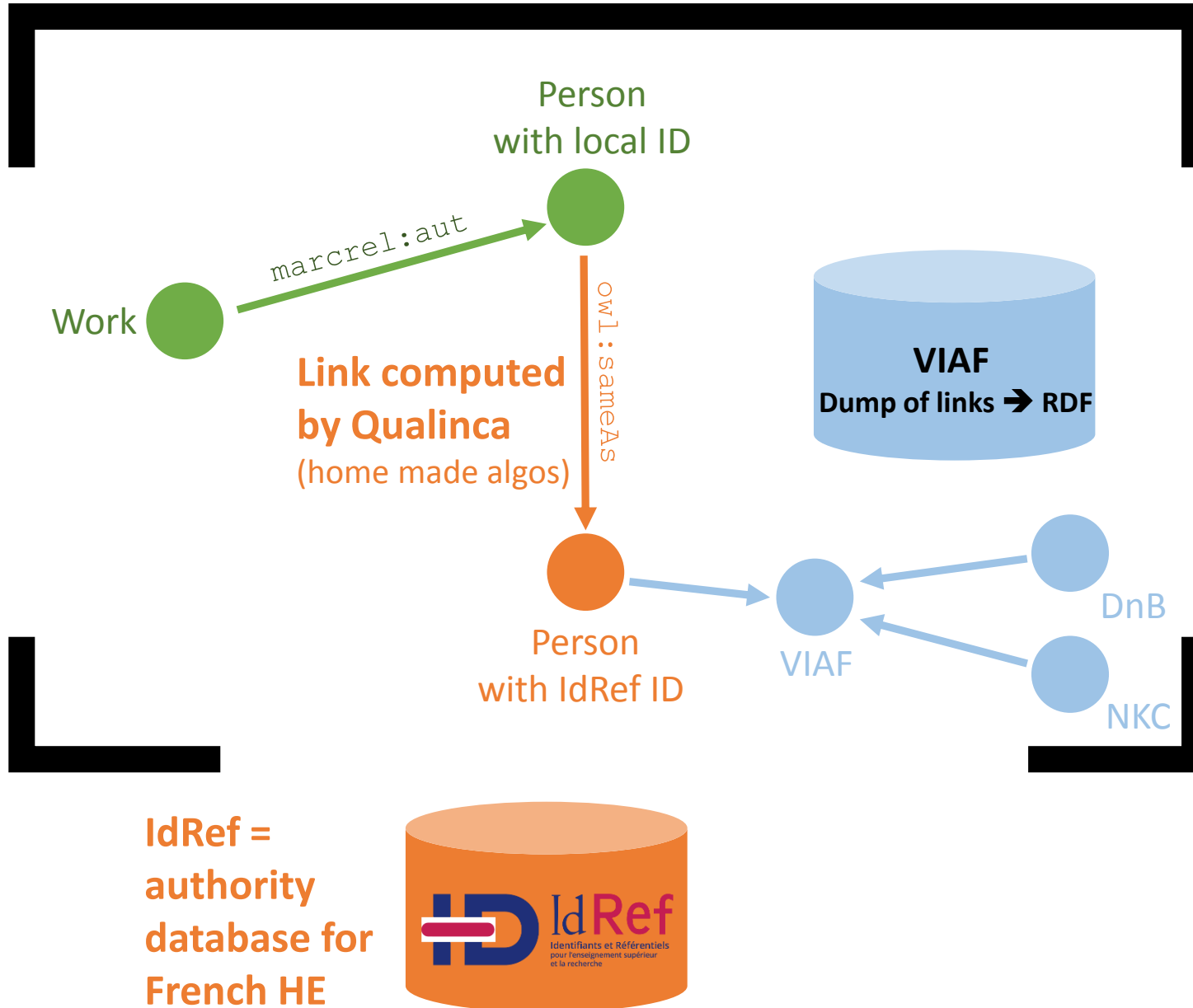
www.bacon.abes.fr



XML

RDF

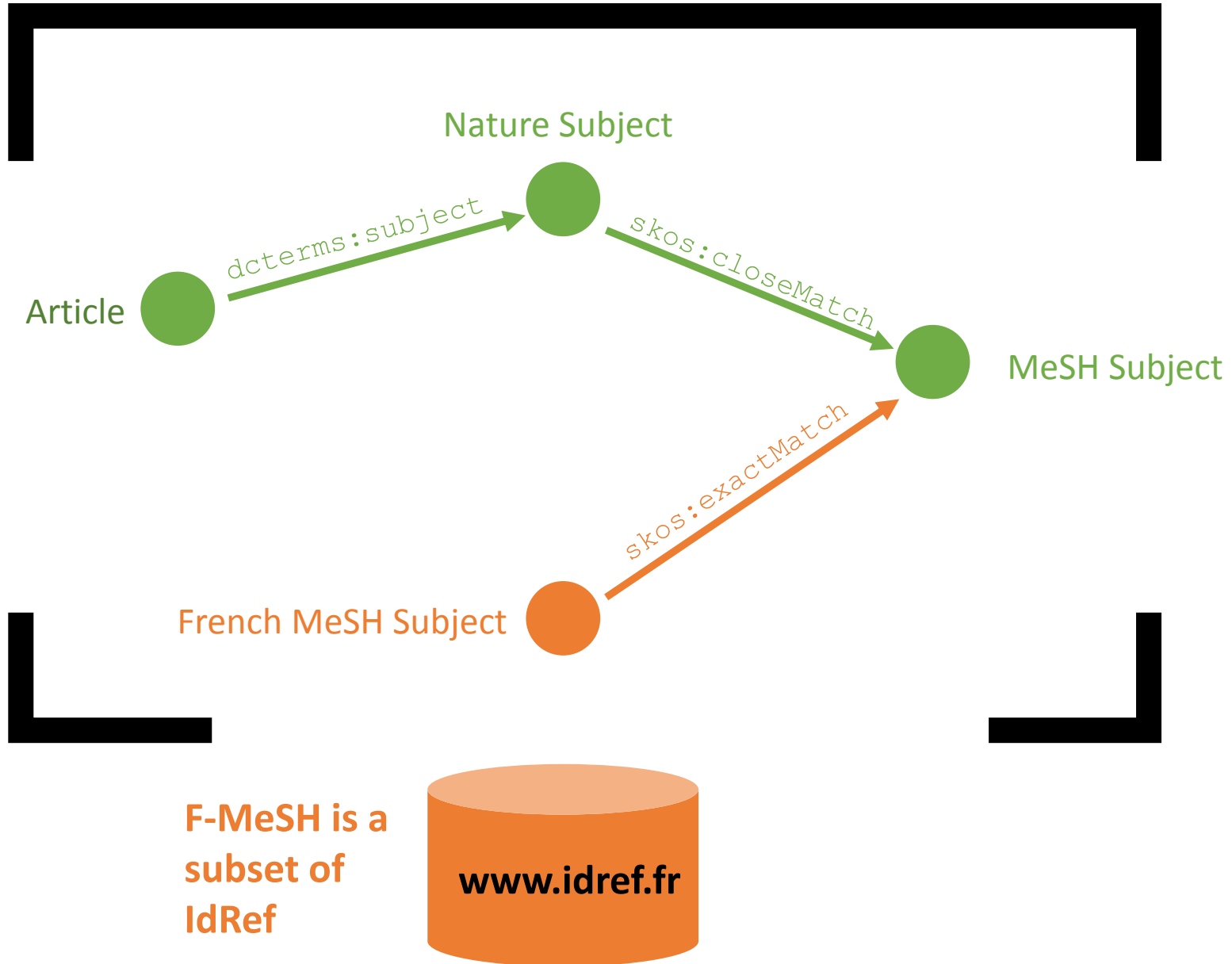
**<XML/>
about each
document**



XML

RDF

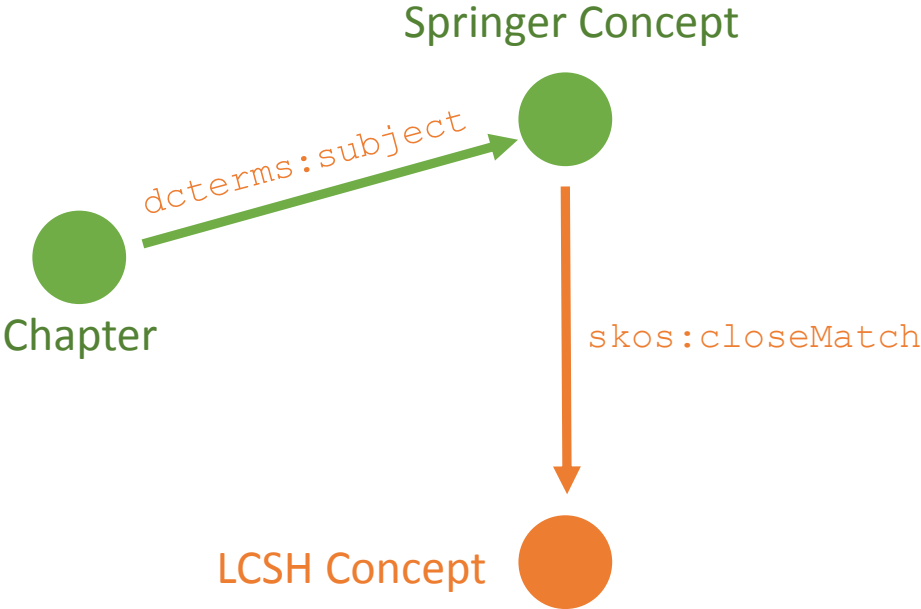
**<XML/>
about each
Nature article**



XML

RDF

<SPRINGER_XML/>
about each
chapter



Springer Concept ID
is the key to merge
the two graphs

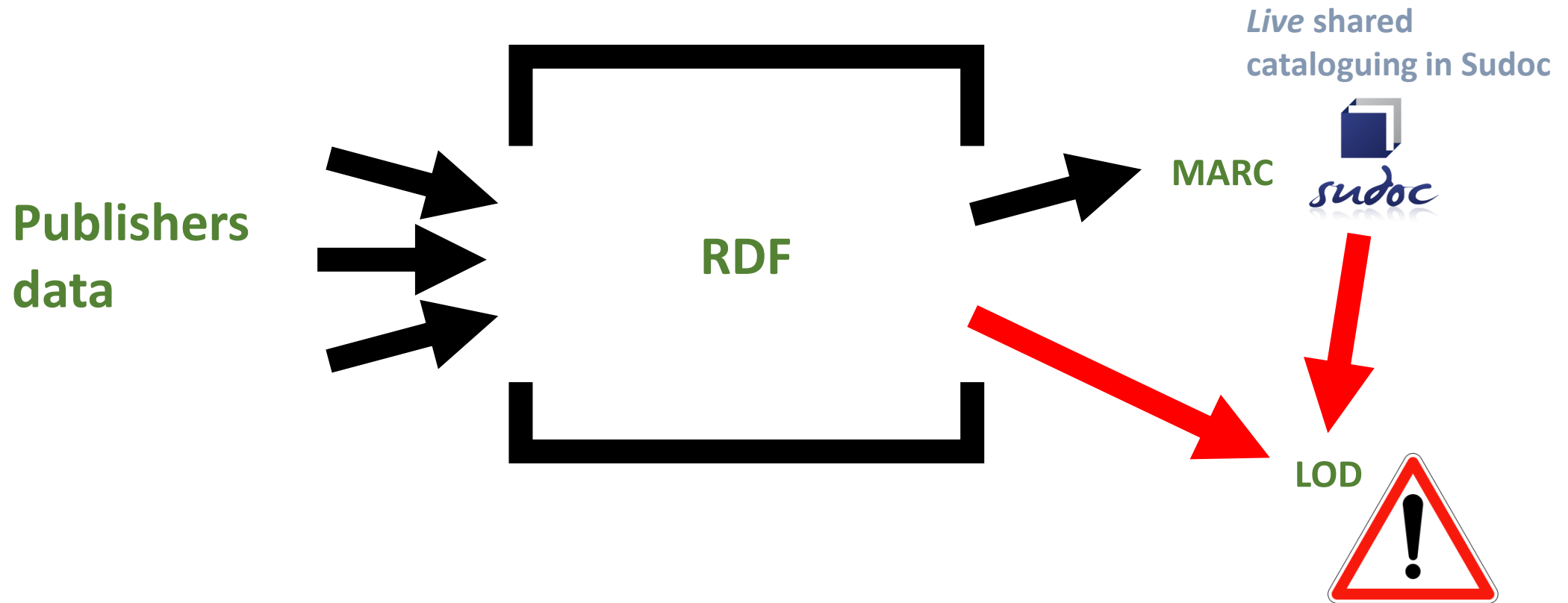
Via Open Refine RDF export

| Springer Concept | LCSH Concept |
|------------------|--------------|
| | |
| Mapping | |

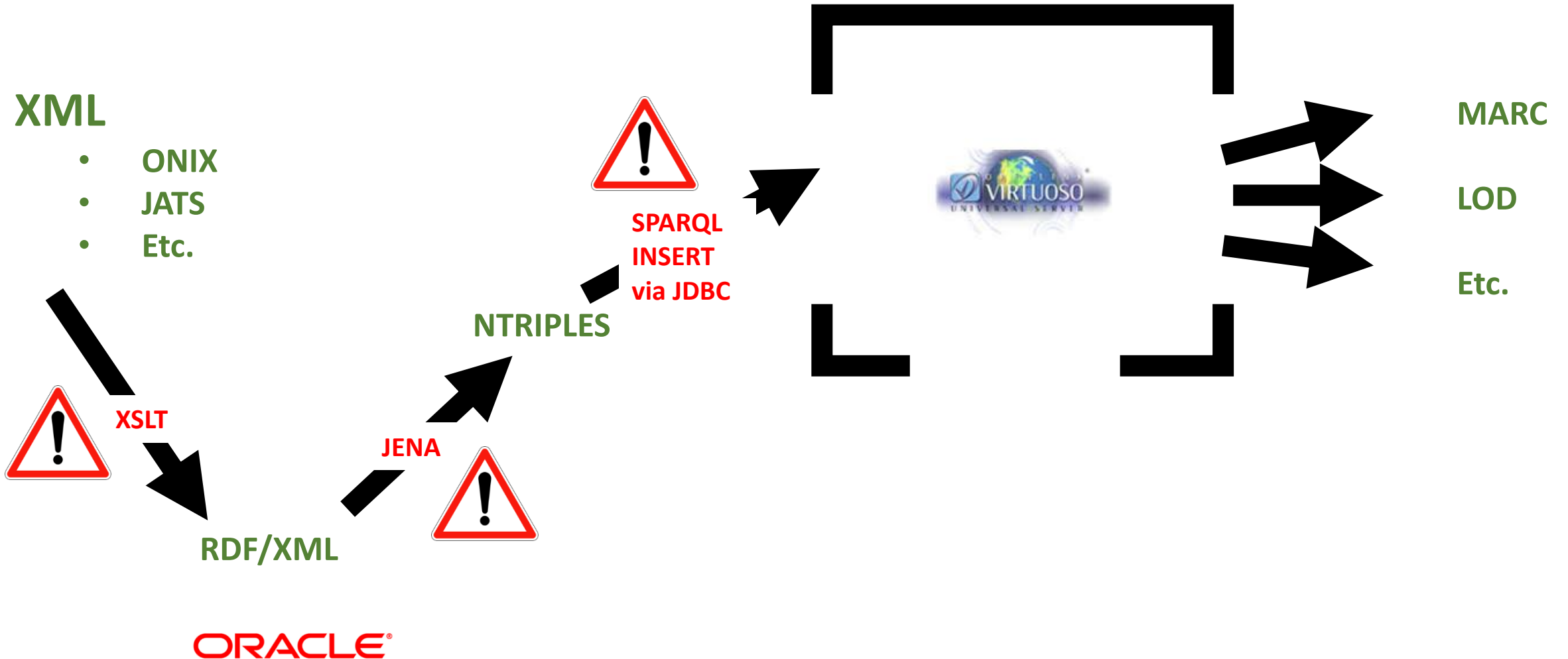


Some concerns...

AS CATALOGERS ARE CATALOGING, SHOULD WE FEED BACK THE RDF BLACK BOX ?

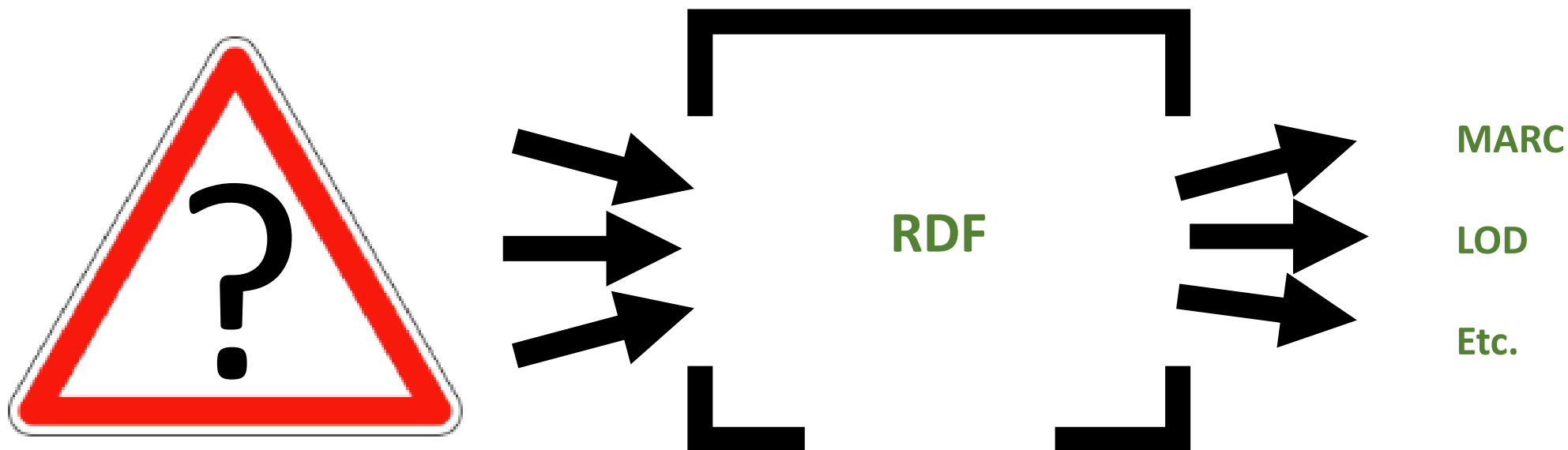


HUMAN and MACHINE BOTTLENECKS



OK...

BUT *WHICH* PUBLISHER DATA PROCESSING ?



NEED FOR A GLOBAL COLLABORATION !

→ NEED FOR A GLOBAL @TODO LIST !

GOK_b ?
Global Open Knowledgebase

Na zdraví !

Questions ?

